

# Educational impact and ethical considerations in using Chatbots in Academia

Dina M. Ibrahim, Njood K. Al-harbi, Amal A. Al-Shargabi

Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia

## Article Info

### Article history:

Received Jun 11, 2024

Revised Sep 13, 2024

Accepted Sep 30, 2024

### Keywords:

Chatbots

ChatGPT-4

Environmental education research

Ethical considerations

Independent learning

Learning tools

Quality of education

## ABSTRACT

Chatbots are getting better every day due to the advancements in their capabilities in today's technological age. This study aims to assess the efficacy of ChatGPT-4 and Gemini in producing scientific articles. Two types of prompts are given: direct questions and complete scenarios. Subsequently, we evaluate the educational and ethical aspects of the produced material by employing statistical analysis. We verify the credibility of references, detect any instances of plagiarism, and ensure the precision of the articles generated by the chatbot. In addition, we utilize topic modeling to assess the extent to which the content of the articles corresponds to the specified topic. According to the findings, Gemini outperformed ChatGPT-4, specifically in scenario questions, where it achieved an accuracy rate of 85%, while ChatGPT-4 only achieved 35% accuracy.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Dina M. Ibrahim

Department of Information Technology, College of Computer, Qassim University

Buraydah, Saudi Arabia

Email: d.hussein@qu.edu.sa

## 1. INTRODUCTION

The present-day era's fast scientific advancement means that people from all walks of life encounter and learn about new technical marvels daily [1]. Who knew that a chatbot could help people save time and work more efficiently? An artificial intelligence chatbot called ChatGPT, which stands for "generative pre-trained transformer," has lately captured the interest of many in the technological community. The artificial intelligence (AI) research and publishing firm OpenAI created it and released it to the public in November 2022. Construction began with OpenAI's GPT-3.5 and GPT-4 large language model (LLM) families, and it was fine-tuned using supervised and reinforcement learning approaches, which is a transfer learning approach [2]. In addition to ChatGPT, another artificial intelligence chatbot called Gemini, previously known as Bard, is designed to simulate conversations with users. Google serves as its power source and it also depends on LLM. Enhanced and streamlined version of LaMDA. It is like ChatGPT, except that it is current and obtains its information directly from the web. Gemini is the official moniker of the new version, which was previously in the development phase. In addition, massive multitask language understanding (MMLU) was the key factor during the development [3].

The emergence of AI technologies, like ChatGPT-4, Gemini, and other chatbots, grants significant capabilities and imposes substantial constraints. Hence, it is imperative to take into account ethical considerations. While it may be advantageous for certain repetitive tasks, such as editing, it presents notable integrity concerns. Furthermore, her rapidity in composing research papers presents a threat to scientific integrity. The present study aims to investigate the influence of chatbots, specifically ChatGPT-4 and Gemini, on academia.

As well as explore methods for educating students and researchers on the ethical and integrity aspects of scientific research. Currently, researchers are investigating the following questions:

1. Can engaging in a conversation with a chatbot generate valuable scientific data for researchers in the academic community? Alternatively, does it give rise to ethical considerations?
2. What are the potential threats associated with chatbots?
3. Does the chatbot's content possess authenticity and dependability?

To address the questions of the study, specific objectives must be achieved, which encompass:

1. Comparative analysis of the functionalities and operations of ChatGPT-4 and Gemini.
2. Conduct experiments with chatbots, including ChatGPT-4 and Gemini.
3. Assess the credibility of the bibliography.
4. Instruct chatbots to produce original articles through two methods: posing direct questions and presenting complete scenarios, followed by conducting plagiarism checks.
5. Examine and analyze the ethical challenges.

Research holds importance in maintaining the integrity of scientific research and in educating students and researchers about the ethical aspects of scientific research. Therefore, it is crucial to employ AI tools responsibly and ethically when conducting academic research and publishing. Moreover, it is crucial to consider copyright, authorship, and accurate citation of information sources. This study investigated the effects of ChatGPT-4 and Gemini to produce original articles through two methods: posing direct questions and presenting complete scenarios, followed by conducting plagiarism checks. While earlier studies have explored the impact of normal questions, they have not explicitly addressed its influence on chatbots, including ChatGPT-4 and Gemini.

The following sections of the paper are structured as follows: section 2 offers a summary of the relevant literature, section 3 outlines the methodologies used for investigation, section 4 presents the results of this research, section 5 provides a discussion of the results, and finally, section 6 presents the conclusion and suggests recommendations for future research.

## 2. RELATED WORKS

This section will provide a brief overview of the literature on ChatGPT, followed by literature on Gemini, and conclude with a discussion on the ethical implications of chatbots in general.

### 2.1. Literature about ChatGPT-4

Zhai [4] predicts ChatGPT will affect all aspects of society. They wrote an academic article to test ChatGPT. The results show that ChatGPT can help researchers write a cogent (partial) report that is correct, insightful, and methodical in two to three hours despite their minimal professional experience. Based on user experience, the author considers ChatGPT and other AI tools' potential effects and proposes changing the learning objectives to teach students how to use them and focus on creativity and critical thinking, which AI tools cannot replace.

Recent research addressed ChatGPT's ethical implications for scientific research, including transparency, bias, informed consent, privacy, responsibility, and integrity. Khlaif [5], concluded that researchers must declare and acknowledge their usage of ChatGPT in their research methods while sticking to research ethics and integrity. In an additional study examining the influence of AI on ethical concerns, specifically in the field of medical publishing practices, Liebreznz *et al.* [6] requested the participation of ChatGPT to provide a commentary on AI and the ethical considerations surrounding medical publishing for lancet digital health. They inquired about how the editorial team can effectively manage the AI-generated academic content, seeking advice from ChatGPT. According to ChatGPT's response to their question, lancet digital health should "carefully consider the ethical implications of publishing articles produced by AI."

In a subsequent publication [7], the author asked ChatGPT, "When streptozotocin-induced diabetes is prepared in growing rats, can you predict its effect on the facial bone growth pattern?" ChatGPT answered. ChatGPT quickly gave a thorough list of references, including their authors, when he asked them. After verifying these referrals, the researcher found them all to be fake. However, the study shows that ChatGPT improved English grammar well. The researcher found that ChatGPT's novel ideas should be validated before publication. Additionally, AI involvement should be stated. Continuing from the last article discussing the authenticity

of ChatGPT references. King [8] uses ChatGPT to complete the essay without any involvement or editing from humans. Regrettably, after a thorough examination, it was discovered that the references requested by the author from ChatGPT were fake.

Ray [9] addresses ChatGPT's history, applications, challenges, bias, ethics, limitations, and future. Data privacy, bias, transparency, autonomy, human agency, emotional manipulation, persuasion, and AI-generated material are among the ethics covered. ChatGPT contributed to scholarly research in linguistically coherent text generation, is grammatically accurate, and has the potential to alter the discipline provided its obstacles and ethical issues are solved.

## 2.2. Literature about Gemini

Bard recently released a new version called Gemini, as a result of the authors' latest research, which analyzed Bard. It is now being developed, and it has been shown that the content provided by Bard in the past is not as accurate when compared to ChatGPT-4 [10]. Nevertheless, the recently launched Gemini Ultra is the initial model to surpass human experts in MMLU. This achievement is accomplished by utilizing a blend of 57 disciplines, including math, physics, history, law, medicine, and ethics, to assess both global knowledge and problem-solving skills, with a score of 90.0% over ChatGPT-4, which has 86.4% [11]. The study's author [12] analyzed three widely used chatbots: ChatGPT, Gemini, and Copilot. The results of the analysis were examined. Most platforms, particularly Google Gemini, have shown a positive response to the personal response prompt. Overall, they indicated that AI chatbots are not dependable for generating ideas.

Imran and Almusharraf [13], stated Google Gemini is useful for education. To minimize fairness issues, its algorithms must be thoroughly evaluated and monitored to ensure they are unbiased and do not prejudice certain learners. Along with other AI chatbots and tools, Gemini must protect data. This requires protecting user data and following data collection ethics. These steps are necessary for AI integration in education. Google Gemini is a major educational technology competitor. Its multimodal, reasoning, and generating skills provide tailored learning, engaging training, and dynamic evaluation. To appropriately utilize this generative AI technology, ethical issues must be considered, responsible advancement must be made, and transparent execution must be ensured. Thus, by prioritizing human-centered design, acknowledging and correcting biases, and maintaining high ethical standards, Google Gemini could create a future where technology allows tailored learning experiences for everyone. The researcher in the study [14], utilizes ChatGPT and Google Gemini as technology resources to develop lesson plans for 7th-grade mathematics, science, literature, and social studies courses. The chatbots were prompted to generate lesson plans tailored to the specified course, subject, and level. The results indicate that the lesson plans generated by both chatbots closely resemble human-authored instructional content in terms of language structures, class activities, and assessments. While the lesson plans designated teachers as facilitators and provided partially productive activities, it was discovered that the technology-integrated activities were severely restricted.

Chatbots in healthcare education are examined in the article [15]. ChatGPT-4 and Gemini are tested on Virology multiple-choice questions (MCQs) in English and Arabic, as well as content quality. ChatGPT-4 and Gemini showed pedagogical promise. ChatGPT-4 had 80% English accuracy and 65% Arabic accuracy, outperforming Gemini's 62.5% and 55%, respectively. Their performance varied across languages, underlining the need for continued work to integrate AI into healthcare education worldwide. The aim of the study introduced by [16], is to determine the proficiency of ChatGPT-4 and Google Gemini in analyzing detailed glaucoma case descriptions and suggesting an accurate surgical plan. ChatGPT-4 demonstrated a commendable level of analysis proficiency when it came to both typical and difficult glaucoma surgical cases. However, Google Gemini exhibited significant constraints in this context, with a high occurrence of inaccurate or unanswered responses.

## 2.3. Literature about ethical aspect

In November 1981, Omni Magazine attributed the authorship of a text to a computer program called Racter [17]. In 1984, Racter's book was published, making it the first book ever created using a computer program [18]. Subsequently, as a result of the development of Racter and AI, copyright issues have garnered significant attention [19]. Can a chatbot be considered an author, considering the rise of AI and chatbots and their extensive use in several industries, particularly in education and student support? The author cited an article [20] that asserted that utilizing an AI text generator without proper acknowledgment would be deemed plagiarism. Plagiarism refers to the unpermitted utilization of someone else's work or ideas without providing proper attribution. According to the authors in the papers [5], [20], it is necessary to mention a source reference

when utilizing AI technologies, regardless of whether the information is generated by humans or AI. Nevertheless, esteemed publications like Nature and Science have explicitly declared that AI chatbots are ineligible to be credited as authors of works published in their journals. The reason for this is that the editorial policies of these publications explicitly indicate that “AI chatbots do not currently meet our criteria for authorship” [21]. AI chatbots cannot function as writers because of their inability to match the present criteria, not because of their lack of human qualities. In the future, AI chatbots could be recognized as authors of scholarly works provided, they fulfill the necessary conditions [17]. Another factor that prevents it from being recognized as an author is its inability to grant authorization for the publication of these papers, which is the basis of the copyright privacy argument [22]. The following ethical issues arise when chatbots create academic content:

- Systematic decision-making errors that yield unfair consequences are called bias. Data gathering, algorithm design, and human interpretation can cause bias. Machine learning models, an artificial intelligence system, can learn and repeat bias patterns in training data, resulting in unfair and biased results [23].
- Human subjects research ethics require informed consent. It involves getting a participant’s permission after presenting them with enough information about the research, its risks, advantages, and alternatives, as well as the chance to ask questions and clarify [24]. However, researchers using AI-generated text lack AI permission. Thus, for transparency and informed consent, researchers must know their techniques to get accurate and reliable results [5].
- Privacy concerns: Chatbots generate text from enormous datasets that may contain sensitive material. As chatbot-generated content becomes more common in scientific research, participant data must be protected. Sharing this information infringes on privacy and creates ethical concerns [5].
- Research integrity, chatbots like Gemini and ChatGPT may provide deceptive information and data. For example, ChatGPT forged references for this work [8]. This shows that chatbots are unreliable and unsuited for use and that researchers are responsible for their research quality and reliability.

In addition, Lund *et al.* [25], examine the impact of GPT/ChatGPT in comparison to other language paradigms, specifically in the context of ethical concerns surrounding technology use in academia, scholarly research, and publishing. GPT3 has demonstrated its versatility, effectiveness, and ability to produce language that closely resembles human speech, rendering it valuable for tasks such as translation, annotation, and question-answering. Moreover, ChatGPT has the capacity to improve search effectiveness and the caliber of scholarly articles. One aspect of the discussion focused on the ethical implications that must be considered, including the ownership of the content. It is currently unclear who holds the rights to the generated text, which raises concerns about copyright issues.

### 3. MATERIALS AND METHODS

The present study methodology is comprised of four distinct experiments. For each experiment, request the chatbot to generate two types of questions: direct questions and scenario questions.

#### 3.1. Direct questions

By posing direct questions or making requests to the chatbots, which are ChatGPT-4 and Gemini, we will inquire without providing further details to observe their respective responses and assess their accuracy. The following are five direct requests for scientific articles. Each article has a distinct title but belongs to the same topic and has the same number of words. The articles should include bibliographies, but it is not necessary to specify whether they should be real or fake.

- Request 1: I want you to write a 1,000-word scientific article on machine learning for financial forecasting and include bibliographies.
- Request 2: I want you to write a 1,000-word scientific article on machine learning for image and video processing and include bibliographies.
- Request 3: I want you to write a 1,000-word scientific article on machine learning for medical diagnosis and include bibliographies.
- Request 4: I want you to write a 1,000-word scientific article on machine learning for natural language processing and include bibliographies.
- Request 5: I want you to write a 1,000-word scientific article on machine learning for social media analysis and include bibliographies.

### 3.2. Scenario questions

In scenario requests, we will ask the chatbots for additional information regarding the title and make the request more accurate by providing data about the requester, such as “I am a master’s student specializing in information technology. I require an in-depth scientific article encompassing an abstract, introduction, literature review, methodology, results, discussion, and a conclusion with future work on the topic about ...”. Furthermore, the article includes the same amount of words in direct questions and pertains to the same topic, but the bibliographies will explicitly indicate the requirement for reliable sources. The following request consists of five scenario inquiries from ChatGPT-4 and Gemini.

- Request 1:

I am a master’s student specializing in information technology. I require an in-depth full scientific article with at least 1,000 words encompassing an abstract, introduction, literature review, methodology, results, discussion, and a conclusion with future work on the topic of machine learning for financial forecasting in stock, and you should include real bibliographies.

- Request 2:

I am a master’s student specializing in information technology. I require an in-depth full scientific article with at least 1,000 words encompassing an abstract, introduction, literature review, methodology, results, discussion, and a conclusion with future work on the topic of machine learning for image and video processing for the detection of faces, and you should include real bibliographies.

- Request 3:

I am a master’s student specializing in information technology. I require an in-depth full scientific article with at least 1,000 words encompassing an abstract, introduction, literature review, methodology, results, discussion, and a conclusion with future work on the topic of machine learning for medical diagnosis for the detection of viruses, and you should include real bibliographies.

- Request 4:

I am a master’s student specializing in information technology. I require an in-depth full scientific article with at least 1,000 words encompassing an abstract, introduction, literature review, methodology, results, discussion, and a conclusion with future work on the topic of machine learning in natural language processing for sentiment analysis of textual data, and you should include real bibliographies.

- Request 5:

I am a master’s student specializing in information technology. I require an in-depth full scientific article with at least 1,000 words encompassing an abstract, introduction, literature review, methodology, results, discussion, and a conclusion with future work on the topic of machine learning in social media advertising analysis, and you should include real bibliographies.

Then, evaluate the results using two criteria: statistical analysis and topic modeling. The statistical analysis comprises the quantification of word count, the total number of bibliographies, the identification of real bibliographies, the identification of fake bibliographies, and the utilization of the plagiarism checker program, Plagiarism Checker X. Then the accuracy will be computed after performing a comparison based on specified criteria presented as (1).

$$Accuracy = \frac{correctedbibliographies}{Thetotalnumberofbibliographies} \quad (1)$$

Topic modeling is an unsupervised machine-learning technique that entails analyzing a large number of documents, articles, feedback, or emails. This study will specifically examine articles generated by chatbots. The main goal of this technique is to detect patterns of words and phrases in the articles, without considering their relation to the topic, but rather focusing on their underlying semantic structure. The provided entity can be characterized as an assemblage of words lacking any particular arrangement or organization [26]. The utilization of latent dirichlet allocation (LDA) for analysis is a computational paradigm that shares similarities with latent semantic analysis (LSA). Nevertheless, LDA and LSA diverge in their methodology for allocating subjects. LDA utilizes word order to allocate topics, intending to comprehend the structure of themes inside documents [26]. Only LDA will be utilized in the current study. Figure 1 provides a comprehensive and complete representation of each experiment.

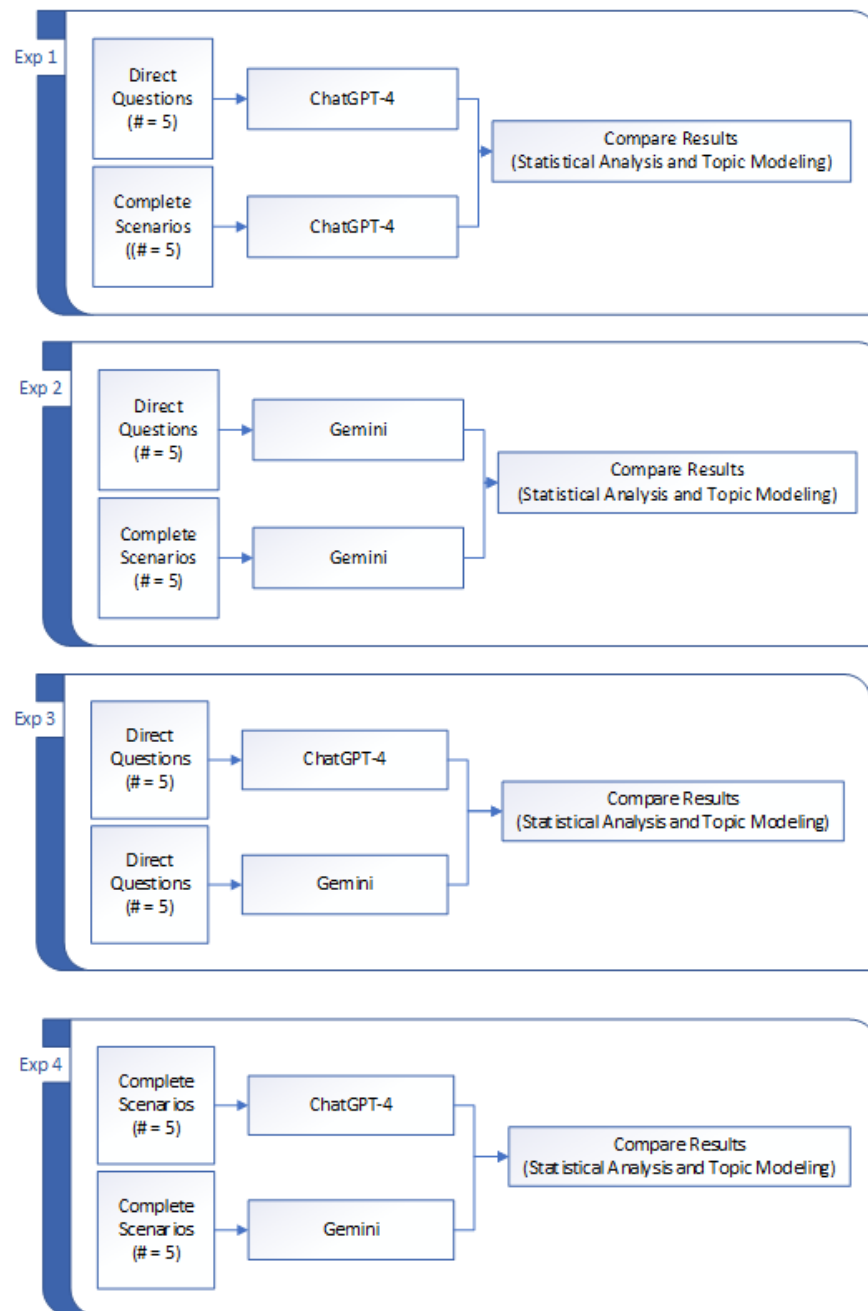


Figure 1. Our experiments steps

## 4. RESULTS AND DISCUSSION

### 4.1. Results

This section provides a detailed comparison between the chatbots ChatGPT-4 and Gemini. Additionally, it discusses the use of a topic modeling technique in the articles listed in Tables 1, 2, 3, and 4, which were subsequently discussed.

#### 4.1.1. ChatGPT-4

When requesting bibliographies without specifying reliable sources in ChatGPT-4, you will receive fake bibliographies, as shown in Table 1. The LDA most relevant terms for five topics and four words for ChatGPT-4 (direct request) are illustrated in Table 2. The topic modeling results for 5 topics and 4 words for ChatGPT-4 (direct request) are shown in Figure 2.

Table 1. Comparison of the academic articles generated by chatbot-4 in the same fields (direct request)

| Article title                                                                                 | Field      | Number of words | The actual word with a bibliography | Number of bibliography | True bibliography | False bibliography | Plagiarism checker X |
|-----------------------------------------------------------------------------------------------|------------|-----------------|-------------------------------------|------------------------|-------------------|--------------------|----------------------|
| 1. Advancements in machine learning for financial forecasting: a comprehensive overview       | Technology | 1,000           | 488                                 | 4                      | 0                 | 4                  | 14%                  |
| 2. Enhancing image and video processing through machine learning: techniques and applications | Technology | 1,000           | 557                                 | 4                      | 0                 | 4                  | 17%                  |
| 3. Machine learning in medical diagnosis: current applications and future prospects           | Technology | 1,000           | 534                                 | 4                      | 0                 | 4                  | 16%                  |
| 4. Advancements in machine learning for natural language processing: a comprehensive review   | Technology | 1,000           | 542                                 | 4                      | 0                 | 4                  | 9%                   |
| 5. Leveraging machine learning for social media analysis: insights and innovations            | Technology | 1,000           | 606                                 | 4                      | 0                 | 4                  | 18%                  |

Table 2. LDA most relevant terms for five topics and four words for ChatGPT-4 (direct request)

| Topic number | Terms       | Rate  |
|--------------|-------------|-------|
| 1            | Application | 0.240 |
|              | Section     | 0.240 |
|              | Efficiency  | 0.040 |
|              | Accuracy    | 0.040 |
| 2            | Section     | 0.171 |
|              | Predictive  | 0.171 |
|              | Efficiency  | 0.171 |
|              | Accuracy    | 0.171 |
| 3            | Section     | 0.067 |
|              | Application | 0.067 |
|              | Efficiency  | 0.067 |
|              | Accuracy    | 0.067 |
| 4            | Ethical     | 0.157 |
|              | Language    | 0.086 |
|              | Implication | 0.086 |
|              | Bias        | 0.086 |
| 5            | Section     | 0.067 |
|              | Application | 0.067 |
|              | Efficiency  | 0.067 |
|              | Accuracy    | 0.067 |

Table 3. Comparison of the academic articles generated by chatbot-4 in the same fields (direct request)

| Article title                                                                                        | Field      | Number of words | The actual word with a bibliography | Number of bibliography | True bibliography | False bibliography        | Plagiarism checker X |
|------------------------------------------------------------------------------------------------------|------------|-----------------|-------------------------------------|------------------------|-------------------|---------------------------|----------------------|
| 1. Enhancing financial forecasting in stock markets through machine learning techniques              | Technology | 1,000           | 568                                 | 5                      | 1                 | 4                         | 14%                  |
| 2. Advancements in machine learning for face detection in image and video processing                 | Technology | 1,000           | 643                                 | 5                      | 3                 | 2 in the publication year | 9%                   |
| 3. Advancing viral detection in clinical diagnostics through machine learning techniques             | Technology | 1,000           | 565                                 | 3                      | 0                 | 3                         | 14%                  |
| 4. Leveraging machine learning in natural language processing for sentiment analysis of textual data | Technology | 1,000           | 588                                 | 4                      | 2                 | 2 in the publication year | 17%                  |
| 5. Enhancing social media advertising through machine learning: an analytical approach               | Technology | 1,000           | 574                                 | 3                      | 1                 | 2                         | 11%                  |

Table 4. LDA most relevant terms for five topics and four words for ChatGPT-4 (scenario request)

| Topic number | Terms          | Rate  |
|--------------|----------------|-------|
| 1            | Result         | 0.015 |
|              | Discussion     | 0.015 |
|              | Base           | 0.015 |
|              | Detection      | 0.015 |
|              | Sentiment      | 0.060 |
| 2            | Base           | 0.033 |
|              | Detection      | 0.033 |
|              | However        | 0.033 |
|              | Model          | 0.044 |
| 3            | Application    | 0.024 |
|              | Machine        | 0.024 |
|              | Practical      | 0.024 |
|              | Discussion     | 0.082 |
| 4            | Dealing        | 0.014 |
|              | Remain         | 0.014 |
|              | Social         | 0.014 |
|              | Trend          | 0.047 |
| 5            | Despite        | 0.047 |
|              | Challenge      | 0.047 |
|              | Dimensionality | 0.047 |

When requesting bibliographies with reliable sources specified in ChatGPT-4, you will receive real and fake bibliographies, as shown in Table 3. LDA most relevant terms for five topics and four words for ChatGPT-4 (scenario request) are illustrated in Table 4. The topic modeling results for 5 topics and 4 words for ChatGPT-4 (scenario request) are shown in Figure 3.



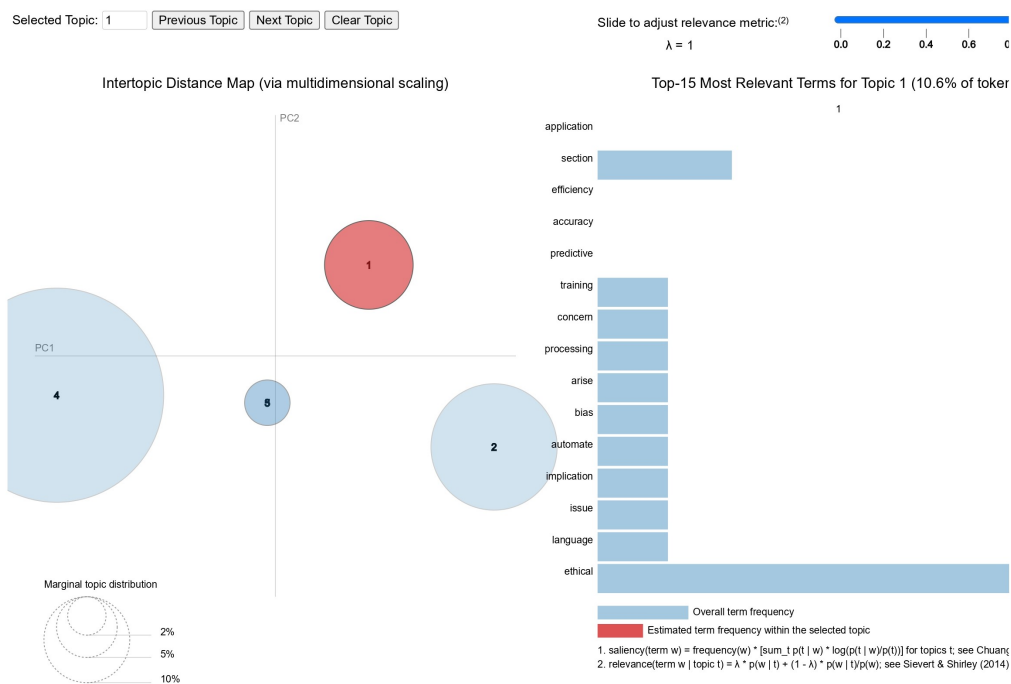


Figure 2. Topic modeling results for 5 topics and 4 words for ChatGPT-4 (direct request)

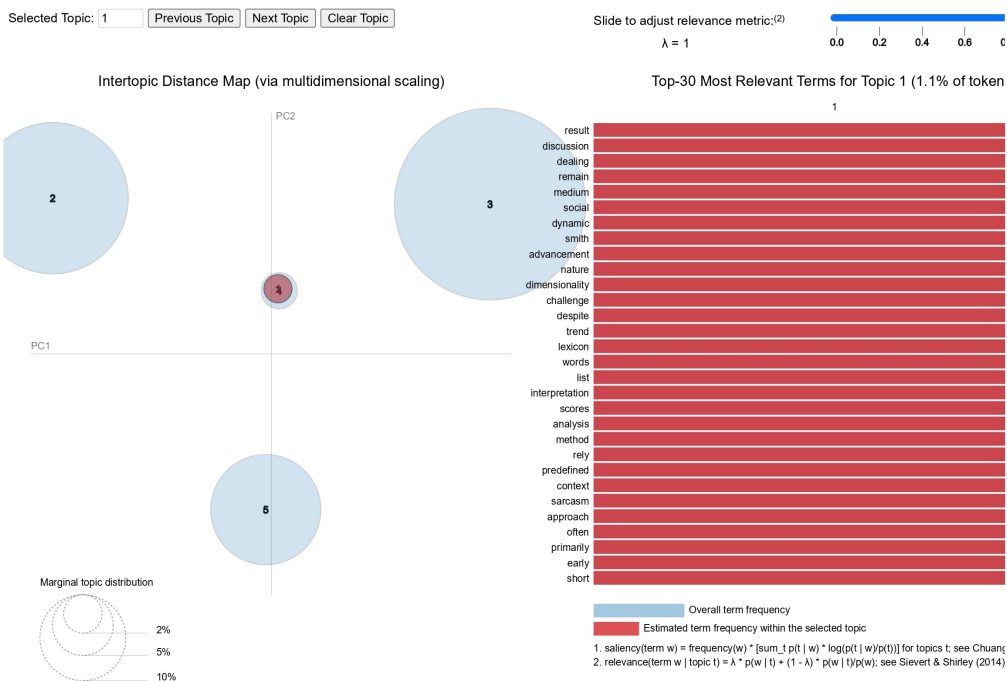


Figure 3. Topic modeling results for 5 topics and 4 words for ChatGPT-4 (scenario request)

#### 4.1.2. Gemini

When requesting bibliographies in Gemini without specifying reliable sources, you may not receive bibliographies for certain articles. These bibliographies can contain a mix of accurate and partially inaccurate information, as indicated in Table 5. The LDA of the most relevant terms for five topics and four words for

Gemini (direct request) is shown in Table 6, while the topic modeling results for 5 topics and 4 words for GEMINI (direct request) are presented in Figure 4.

When requesting bibliographies specifying reliable sources in Gemini, you will receive real and fake bibliographies as shown in Table 7. The LDA of the most relevant terms for five topics and four words for Gemini (direct request) is shown in Table 8, while the topic modeling results for 5 topics and 4 words for GEMINI (direct request) are presented in Figure 5.

Table 5. Comparison of the academic articles generated by Gemini in the same fields (direct request)

| Article title                                                                                         | Field      | Number of words | The actual word with a bibliography | Number of bibliography | True bibliography | False bibliography            | Plagiarism checker X |
|-------------------------------------------------------------------------------------------------------|------------|-----------------|-------------------------------------|------------------------|-------------------|-------------------------------|----------------------|
| 1. Machine learning for financial forecasting: unveiling hidden patterns and predicting market trends | Technology | 1,000           | 1,587                               | 7                      | 7                 | 1 & 1 in the publication year | 19%                  |
| 2. Unveiling the potential: machine learning for image and video processing                           | Technology | 1,000           | 1,180                               | 7                      | 7                 | 0                             | 22%                  |
| 3. Machine learning for medical diagnosis: revolutionizing healthcare through pattern recognition     | Technology | 1,000           | 1,121                               | 8                      | 6                 | 2                             | 19%                  |
| 4. Unveiling the power of language: machine learning for natural language processing                  | Technology | 1,000           | 1,333                               | 6                      | 4                 | 1 & 1 in the publication year | 27%                  |
| 5. Unveiling the social pulse: machine learning for social media analysis                             | Technology | 1,000           | 1,054                               | 5                      | 4                 | 1                             | 16%                  |

Table 6. LDA most relevant terms for five topics and four words for Gemini (direct request)

| Topic number | Terms       | Rate  |
|--------------|-------------|-------|
| 1            | Event       | 0.165 |
|              | Black       | 0.113 |
|              | Market      | 0.113 |
|              | Volatile    | 0.062 |
| 2            | Speech      | 0.143 |
|              | Convert     | 0.143 |
|              | Speak       | 0.143 |
|              | Recognition | 0.143 |
| 3            | Recognition | 0.059 |
|              | Convert     | 0.059 |
|              | Language    | 0.059 |
|              | Speak       | 0.059 |
| 4            | Language    | 0.059 |
|              | Recognition | 0.059 |
|              | Speak       | 0.059 |
|              | Speech      | 0.059 |
| 5            | Speak       | 0.059 |
|              | Recognition | 0.059 |
|              | Language    | 0.059 |
|              | Convert     | 0.059 |

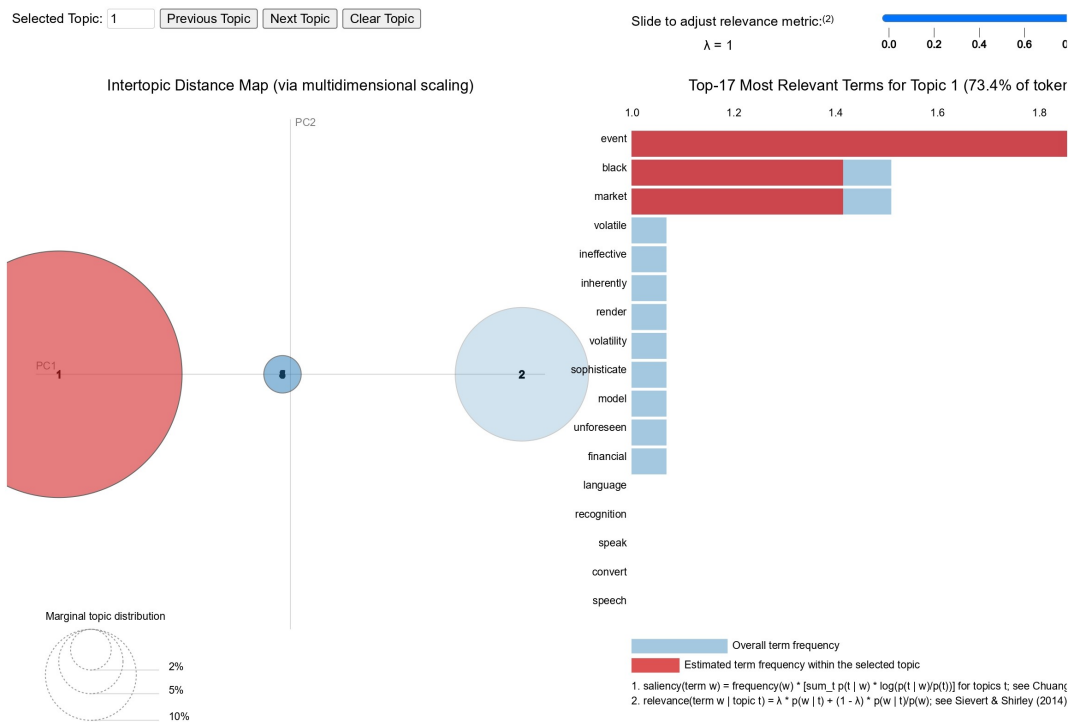


Figure 4. Topic modeling results for 5 topics and 4 words for GEMINI (direct request)

Table 7. Comparison of the academic articles generated by Gemini in the same fields (scenario request)

| Article title                                                                          | Field      | Number of words | The actual word with a bibliography | Number of bibliography | True bibliography | False bibliography                                   | Plagiarism checker X |
|----------------------------------------------------------------------------------------|------------|-----------------|-------------------------------------|------------------------|-------------------|------------------------------------------------------|----------------------|
| 1. Machine learning for stock price forecasting: a critical review and exploration     | Technology | 1,000           | 1,401                               | 3                      | 1                 | 1 & 1 in the publication year                        | 10%                  |
| 2. Machine learning for face detection in images and videos: a comprehensive survey    | Technology | 1,000           | 1,128                               | 7                      | 7                 | 0                                                    | 22%                  |
| 3. Machine learning for medical diagnosis: a powerful tool for viral disease detection | Technology | 1,000           | 1,016                               | 3                      | 3                 | But there is a mistake in some of the author's names | 23%                  |
| 4. Machine learning for sentiment analysis: unveiling emotions in textual data         | Technology | 1,000           | 1,333                               | 5                      | 4                 | 1 in the publication year                            | 26%                  |
| 5. Unveiling consumer insights: machine learning for social media advertising analysis | Technology | 1,000           | 1,440                               | 2                      | 2                 | 0                                                    | 14%                  |

Table 8. LDA most relevant terms for five topics and four words for Gemini (scenario request)

| Topic number | Terms       | Rate  |
|--------------|-------------|-------|
| 1            | Infection   | 0.037 |
|              | Detecting   | 0.037 |
|              | Sensitivity | 0.037 |
|              | Always      | 0.037 |
| 2            | Infection   | 0.037 |
|              | Detecting   | 0.037 |
|              | Always      | 0.037 |
|              | Early       | 0.037 |
| 3            | Training    | 0.077 |
|              | Model       | 0.077 |
|              | Prediction  | 0.042 |
|              | Portion     | 0.042 |
| 4            | Limited     | 0.105 |
|              | Sensitivity | 0.105 |
|              | Always      | 0.105 |
|              | Early       | 0.105 |
| 5            | Detecting   | 0.037 |
|              | Infection   | 0.037 |
|              | Early       | 0.037 |
|              | Always      | 0.037 |

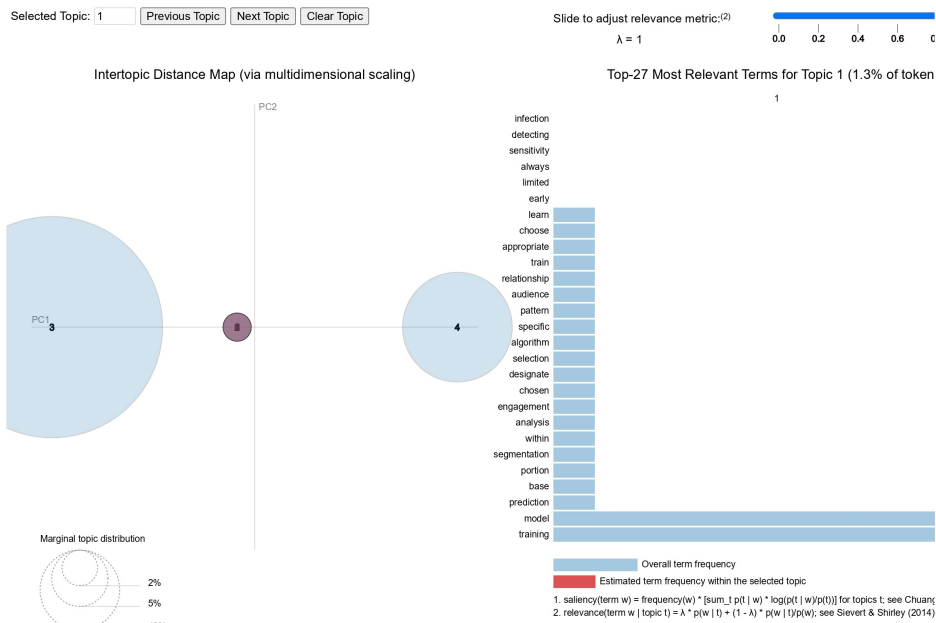


Figure 5. Topic modeling results for 5 topics and 4 words for GEMINI (scenario request)

## 4.2. Discussions

This section will analyze the outcomes of the Chatbots articles in both direct questions vs scenario questions in ChatGPT-4 and Gemini, compare the outcomes of Topic modeling and ethical aspects, and provide a summary of the findings.

### 4.2.1. Comparison of the Chatbots articles outcomes from Tables 1, 3, 5, and 7

An analysis will be conducted to compare the accuracy of the initial ChatGPT-4 chatbot when presented with direct questions versus when provided with a comprehensive scenario prompt. The Gemini chatbot will also be evaluated using the same comparison criteria. Ultimately, we will exclusively consider the direct questions or scenario prompts that are superior in both chatbots, to conduct a thorough comparison between ChatGPT-4 and Gemini.

- Direct questions vs scenario questions (ChatGPT-4)

According to the results presented in Table 1, if the user requests direct questions to the chatbot to include bibliographies without stating that they should be real, the outcome is that the 5 articles contain a total of 20 references, all of which are fake. Beyond the minimum word count requirement of 1000, the maximum number of words received is 606, which falls short of the requirement. Additionally, the plagiarism rate did not exceed 18%, which is satisfactory. Therefore, the accuracy of ChatGPT-4 in answering direct questions is 0%.

On the other hand, Table 2 presents the result if the user requests scenario questions to the chatbot to include bibliographies stating that they should be real, the outcome is that the 5 articles contain a total of 20 references, the majority of the references, specifically 13 out of 20, are deemed to be false, 4 in the publication year and 6 are fully faked. Conversely, the remaining 7 references are considered to be real, resulting in an accuracy rate of 35.00%. This accuracy rate surpasses that of the direct questions presented in Table 1. In addition to the minimum word count of 1000, the maximum number of words received is 634, which is insufficient to meet the requirement. In addition, the rate of plagiarism did not surpass 17%, and both have I. made slight improvements to Table 1.

- Direct questions vs scenario questions (Gemini)

According to Table 4 when a user requests a direct question to the chatbot without specifying reliable references. The result is that out of the five papers, which collectively have thirty-three references, 26 of them are real, 2 have incorrect publication years and 5 are fake, with an accuracy of 78.78%. The minimum word count is 1054, while the maximum word limit is 1587. It is evident that Gemini surpasses ChatGPT-4 in terms of meeting the requirements of accuracy and word count. However, when it comes to the rate of plagiarism, it surpasses the rate of ChatGPT-4, with the highest recorded plagiarism rate being 27%.

Whereas, when the user requests the chatbot for scenario questions that include bibliographies and specifies that they should be real, Table 6 displays the findings. It reveals that the five articles have a combined total of twenty references. Out of these, seventeen references are real, while 2 references have incorrect publication years, and just 1 reference is completely faked. The accuracy is 85.00%. The received word count ranges from a minimum of 1016 to a maximum of 1440, meeting the requirements stated. The plagiarism rate ranges from a minimum of 10% to a maximum of 26%.

- Direct questions vs scenario questions (ChatGPT-4 and Gemini)

The Gemini chatbot is considered superior due to its impressive accuracy ratings of 78.78% for direct questions and 85.00% for scenario questions. Conversely, ChatGPT-4 receives 0% and 35.00%, respectively, and does not meet the other requirements. To recap, Gemini in Scenario Questions proved to be the most precise method for obtaining academic responses, achieving an accuracy rate of 85%.

#### 4.2.2. Comparison of the outcomes of the Topic modeling from Tables 2, 4, 6, and 8

- Direct questions vs scenario questions (ChatGPT-4)

The direct questions are categorized into five topics, and their word rates are provided in Table 2. Figure 2 shows Topic 1, which is associated with 10.6% of tokens, encompassing certain phrases and their corresponding rate. Figure 6 shows Topic 2, which is associated with 21.6% of tokens and also shows Topic 3 which is associated with 2.8% of tokens. Figure 7 shows Topic 4, which is associated with 62.1% of tokens. Lastly, the figure shows Topic 5 associated with 2.8% of tokens. Two topics, specifically Topics 3 and 5 in Figure 7, are clustered together.

In the scenario, the technique for handling questions is identical to that of direct questions in topic modeling, as shown in Table 4. Figure 3 shows Topic 1, which is associated with 1.1% of tokens, encompassing certain phrases and their corresponding rate. Figure 8 shows Topic 2, which is associated with 31% of tokens and also shows Topic 3, which is associated with 49.7% of tokens. Figure 9 shows Topic 4, which is associated with 1.8% of tokens. Lastly, Topic 5 is represented by Figure 13, which is associated with 16.5% of tokens. Two topics, specifically Topic 1 and Topic 4 in Figure 9, are clustered together.

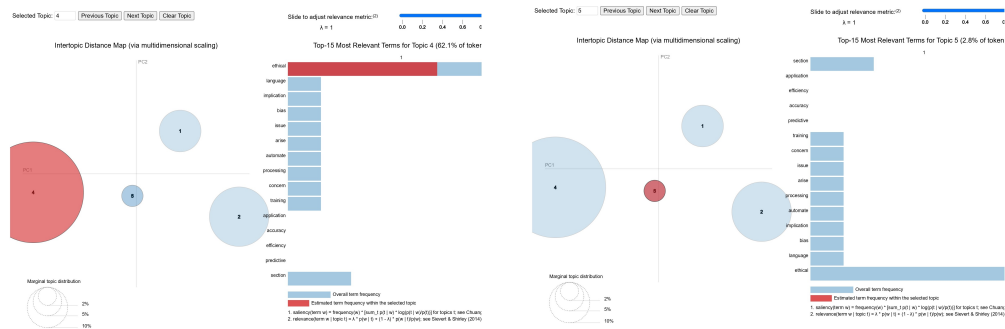


Figure 6. Topic 2 and 3 ChatGPT-4 (direct request)

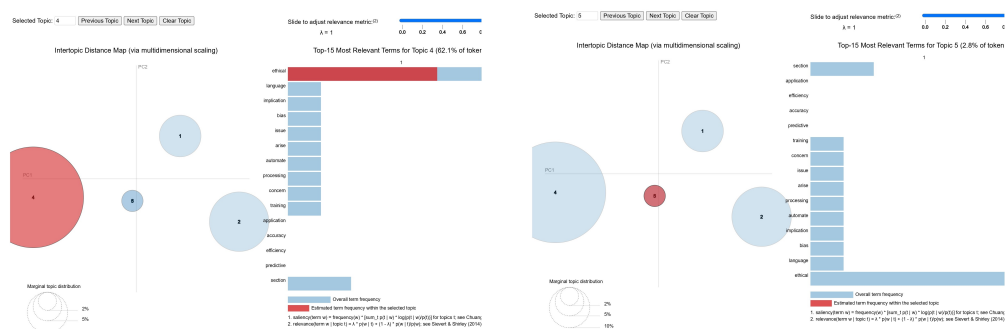


Figure 7. Topic 4 and 5 ChatGPT-4 (direct request)

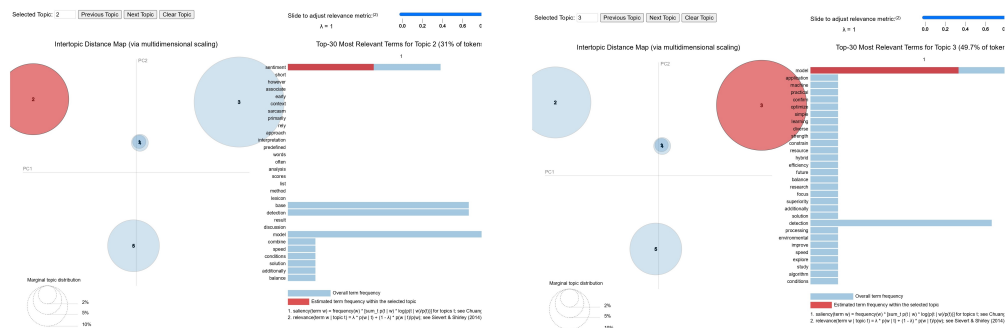


Figure 8. Topic 2 and 3 ChatGPT-4 (scenario request)

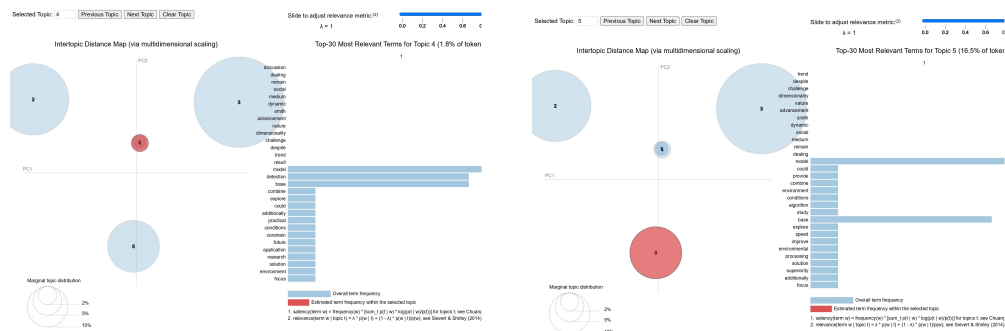


Figure 9. Topic 4 and 5 ChatGPT-4 (scenario request)

### - Direct questions vs. scenario questions (Gemini)

The identical procedure was carried out on the Gemini chatbot. The direct questions in Table 6 illustrate the relevant words to each Topic. Figure 10 shows Topic 2 which is associated with 21.5% of tokens. Also shows Topic 3, which is associated with 1.7% of tokens. Figure 11 shows Topic 4, which is associated with 1.7% of tokens. Lastly, the figure shows Topic 5, which is associated with 1.7% of tokens. Three topics, specifically Topic 3 in Figure 10, Topic 4 and Topic 5 in Figure 11 are clustered together.

The scenario question in Table 8 illustrates the relevant words to each Topic. Figure 5 shows Topic 1, associated with 1.3% of tokens. Figure 12 shows Topic 2, associated with 1.3% of tokens. the figure also shows Topic 3, which is associated with 77.3% of tokens. Figure 13 shows Topic 4 associated with 19% of tokens. Lastly, Topic 5 is represented by Figure 13, which is associated with 1.3% of tokens. Three topics, specifically Topic 1, Topic 2 in Figure 12, and Topic 5 in Figure 13 are clustered together.

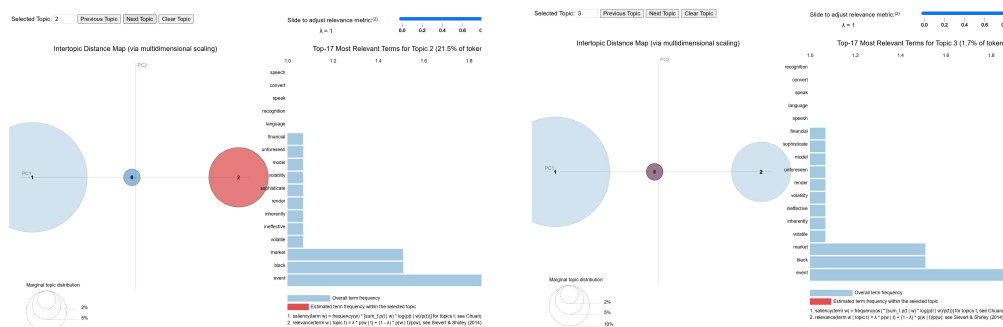


Figure 10. Topic 2 and 3 Gemini (direct request)

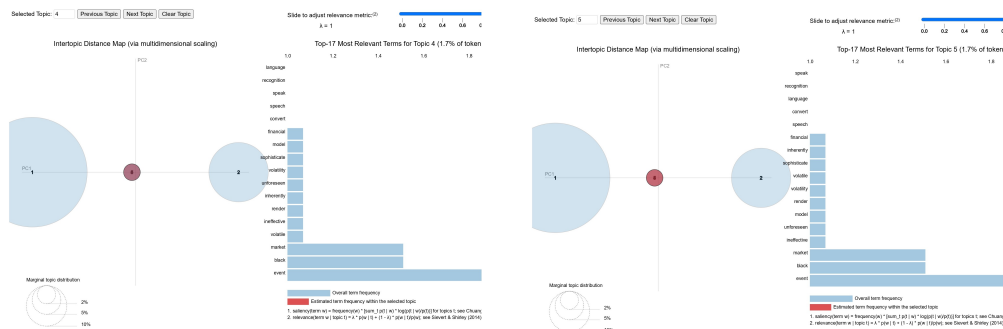


Figure 11. Topic 4 and 5 Gemini (direct request)

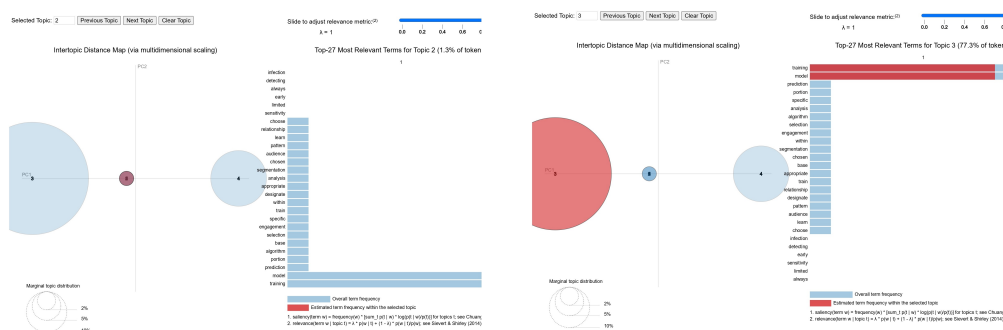


Figure 12. Topic 2 and 3 Gemini (scenario request)

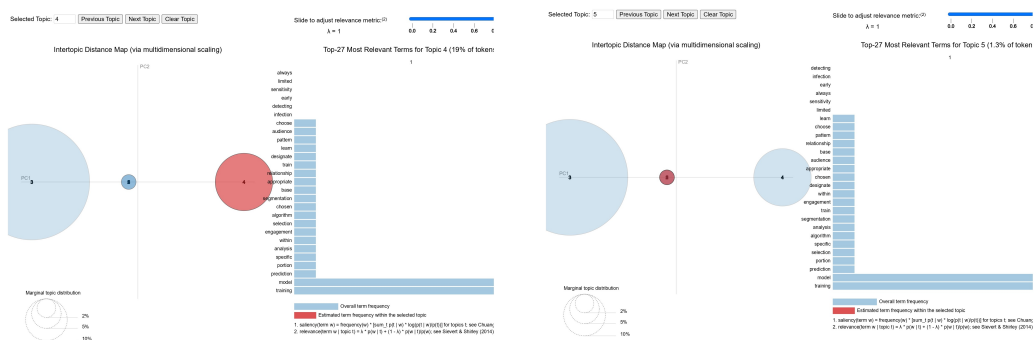


Figure 13. Topic 4 and 5 Gemini (scenario request)

#### - Direct questions vs scenario questions (ChatGPT-4 and Gemini)

To recap, the above Figures demonstrate that Gemini has clustered three topics whereas ChatGPT-4 was clustered two. This indicates that Gemini is more relevant to the selected topics.

**Ethical aspect:** Chatbots cannot be regarded as authors of scientific publications due to their lack of absolute correctness. The scientific study relies on humans for legitimacy and integrity, and a chatbot cannot be considered a human entity. It can be utilized for laborious repetitive jobs that are in line with earlier articles [25], but in academic research, it lacks explicit information and modifies its responses if you express disagreement. Unlike chatbots, researchers possess a comprehensive understanding of their research topic and can justify their conclusions. However, the success of their research is contingent upon their ethics and integrity. Furthermore, it is worth noting that there are currently no chatbots available that can provide 100% accurate bibliographies. The accuracy of bibliographic information is crucial for ensuring the correctness of articles. It is not acceptable to make any mistakes in the content, author's name, or publication year. Additionally, the chatbot's content appears to lack specificity and does not demonstrate the level of expertise one would anticipate from a master's student. It also lacks comprehensiveness.

#### Summary of findings:

1. The utilization of direct questions versus scenario questions on ChatGPT-4 is unsatisfactory. There has been improvement in the scenario question, but it is not significant.
2. The performance of direct questions and scenario questions on Gemini is superior to that of ChatGPT-4, particularly the scenario questions, which yielded better outcomes with 85% accuracy.
3. In topic modeling, both direct questions and scenario questions in ChatGPT4 are clustered into two topics, however, in Gemini, they are clustered into three topics. This indicates that the content generated by Gemini is more relevant to the topics compared to ChatGPT-4.
4. No chatbot achieves 100% accuracy in generating precise bibliographies.

## 5. CONCLUSIONS AND FUTURE WORK

Chatbots are growing rapidly, and their use in higher education can improve text summarization, creation, and translation. These chatbots raise concerns about academic honesty and plagiarism. These chatbots may increase fraud. It can be hard to tell automated text from human-written material. This study employed two methods to evaluate the content produced by chatbots, including direct and scenario questions, to assess the chatbot's ability in academic writing.

Based on the results of this analysis, the use of two chatbots (ChatGBT-4 and Gemini) through topic modeling was found to be relevant to the topic. However, it cannot be considered scholarly or suitable for academic research because it did not meet several criteria assessed in this investigation. The Gemini chatbot is regarded as outstanding since it achieves impressive accuracy rates of 78.78% for direct questions and 85.00% for scenario questions. In contrast, ChatGPT-4 earns a score of 0% and 35.00% for the mentioned criteria, but it fails to fulfill the remaining standards. In summary, Gemini demonstrated the highest level of accuracy, with an 85% success rate, when it came to obtaining academic solutions in scenario questions.



Moreover, it is important to mention that there are presently no chatbots accessible that can deliver bibliographies with complete accuracy. Ensuring the accuracy of bibliographic information is essential for guaranteeing the correctness of articles. Errors in the content, author's name, or publication year are not permissible. Additionally, the chatbot's information seems to lack precision and does not exhibit the level of proficiency that is typically associated with a master's student. Furthermore, it lacks comprehensiveness. In future research, we will evaluate various chatbots, such as Microsoft's Copilot.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge Qassim University, represented by the Deanship of Scientific Research, on the financial support for this research under the number (2023-SDG-1-BSRC36790) during the academic year 1445 AH / 2023 AD.

## REFERENCES

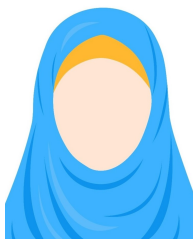
- [1] M. S. Rahaman, M. M. T. Ahsan, N. Anjum, M. M. Rahman, and M. N. Rahman, "The AI race is on! Google's bard and openai's Chatgpt head to head: an opinion article," *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4351785.
- [2] openai.com, "Introducing ChatGPT," *Openai.Com*, 2022. <https://openai.com/about> (accessed Jun. 06, 2024).
- [3] S. Hsiao, "Get more done with Gemini: Try 1.5 Pro and more intelligent features," *Google*, 2024, [Online]. Available: <https://blog.google/products/gemini/google-gemini-update-may-2024/>.
- [4] X. Zhai, "ChatGPT user experience: implications for education," *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4312418.
- [5] Z. N. Khlaif, "Ethical concerns about using AI-generated text in scientific research," *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4387984.
- [6] M. Liebrez, R. Schleifer, A. Buadze, D. Bhugra, and A. Smith, "Generating scholarly content with ChatGPT: ethical challenges for medical publishing," *The Lancet Digital Health*, vol. 5, no. 3, pp. e105–e106, Mar. 2023, doi: 10.1016/S2589-7500(23)00019-5.
- [7] S. G. Kim, "Using ChatGPT for language editing in scientific articles," *Maxillofacial Plastic and Reconstructive Surgery*, vol. 45, no. 1, p. 13, Mar. 2023, doi: 10.1186/s40902-023-00381-x.
- [8] M. R. King, "A conversation on artificial intelligence, chatbots, and plagiarism in higher education," *Cellular and Molecular Bioengineering*, vol. 16, no. 1, pp. 1–2, Feb. 2023, doi: 10.1007/s12195-022-00754-8.
- [9] P. P. Ray, "ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope," *Internet of Things and Cyber-Physical Systems*, vol. 3, pp. 121–154, 2023, doi: 10.1016/j.iotcps.2023.04.003.
- [10] N. K. Al-Harbi and A. A. Al-Shargabi, "An exploratory analysis of using Chatbots in Academia," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 12, pp. 122–132, 2023, doi: 10.14569/IJACSA.2023.0141212.
- [11] S. Pichai and D. Hassabis, "Introducing Gemini: our largest and most capable AI model," *Google*, 2023. <https://blog.google/technology/ai/google-gemini-ai/> (accessed Jun. 06, 2024).
- [12] C. Wilson, D. Collections, B. Hawley, and N. Hollans, "Using AI Chatbots as ideation machines," 2024. [https://digitalcollections.lipscomb.edu/eml\\_student/9](https://digitalcollections.lipscomb.edu/eml_student/9).
- [13] M. Imran and N. Almusharraf, "Google Gemini as a next generation AI educational tool: a review of emerging educational technology," *Smart Learning Environments*, vol. 11, no. 1, p. 22, May 2024, doi: 10.1186/s40561-024-00310-z.
- [14] A. Baytak, "The content analysis of the lesson plans created by ChatGPT and Google Gemini," *Research in Social Sciences and Technology*, vol. 9, no. 1, pp. 329–350, Mar. 2024, doi: 10.46303/ressat.2024.19.
- [15] M. Sallam et al., "The performance of OpenAI ChatGPT-4 and Google Gemini in virology multiple-choice questions: a comparative analysis of English and Arabic responses," *BMC Research Notes*, vol. 17, no. 1, p. 247, Sep. 2024, doi: 10.1186/s13104-024-06920-7.
- [16] M. M. Carlà et al., "Large language models as assistance for glaucoma surgical cases: a ChatGPT vs. Google Gemini comparison," *Graefes' Archive for Clinical and Experimental Ophthalmology*, vol. 262, no. 9, pp. 2945–2959, Sep. 2024, doi: 10.1007/s00417-024-06470-5.
- [17] J. Y. Lee, "Can an artificial intelligence chatbot be the author of a scholarly article?," *Journal of Educational Evaluation for Health Professions*, vol. 20, p. 6, Feb. 2023, doi: 10.3352/jeehp.2023.20.6.
- [18] "The first book written by a computer program," *Normans, Jeremy*, 1984. <https://www.historyofinformation.com/detail.php?id=3351> (accessed May 01, 2023).
- [19] T. L. Butler, "Can a computer be an author? Copyright aspects of artificial intelligence," *Hastings Communications Entertainment Law Journal*, vol. 4, no. 4, pp. 707–747, 1981.
- [20] B. L. Frye, "Should using an AI text generator to produce academic writing be plagiarism? Be lagiarism? Chat GPT," *Fordham Intellectual Property, Media and Entertainment Law Journal*, vol. 33, no. 4, p. 2023, 2022, [Online]. Available: <https://ir.lawnet.fordham.edu/iplj/vol33/iss4/5>.
- [21] Springer Nature, "Authorship," Springer Nature, [Online]. Available: <https://www.nature.com/nature-portfolio/editorial-policies/authorship>, (Accessed: Jun. 06, 2024).
- [22] C. Stokel-Walker, "ChatGPT listed as author on research papers: many scientists disapprove," *Nature*, vol. 613, no. 7945, pp. 620–621, Jan. 2023, doi: 10.1038/d41586-023-00107-z.
- [23] E. Ferrara, "Fairness and bias in artificial intelligence: a brief survey of sources, impacts, and mitigation strategies," *Sci*, vol. 6, no. 1, p. 3, Dec. 2024, doi: 10.3390/sci6010003.

- [24] S. Manti and A. Licari, "How to obtain informed consent for research," *Breathe*, vol. 14, no. 2, pp. 145–152, Jun. 2018, doi: 10.1183/20734735.001918.
- [25] B. D. Lund, T. Wang, N. R. Mannuru, B. Nie, S. Shimray, and Z. Wang, "ChatGPT and a new academic reality: artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing," *Journal of the Association for Information Science and Technology*, vol. 74, no. 5, pp. 570–581, May 2023, doi: 10.1002/asi.24750.
- [26] F. Pascual, "Introduction to topic modeling," no. September. SAGE Publications Ltd, London United Kingdom, pp. 1–42, 2019, doi: 10.4135/9781526499776.

## BIOGRAPHIES OF AUTHORS



**Dina M. Ibrahim**    has been an associate professor at the Information Technology Department, College of Computer, Qassim University, KSA, since 2015. In addition, Dina works as an assistant professor at the Computers and Control Engineering Department, Faculty of Engineering, Tanta University, Egypt. She was born in the United Arab Emirates, and her B.Sc., M.Sc., and Ph.D. degrees were obtained from the Computers and Control Engineering Department, Faculty of Engineering, Tanta University, in 2002, 2008, and 2014, respectively. Dina worked as a consultant engineer, database administrator, and vice manager on the Management Information Systems (MIS) Project, at Tanta University, Egypt, from 2008 until 2014. Her research interests include networking, wireless communications, machine learning, security, and the Internet of Things. Dina has published about 70 articles in various refereed international journals and conferences. She has been serving as a reviewer in the Wireless Network (WINE) Journal since 2015. Dina has also served as a co-chair of the International Technical Committee for the Middle East Region of the ICCMIT conference since 2020. She can be contacted at email: d.hussein@qu.edu.sa and dina.mahmoud@f-eng.tanta.edu.eg.



**Njood K. Al-harbi**    she received a B.Sc. degree in information technology from Qassim University in Saudi Arabia. Currently, Njood studying for a M.Sc. degree in information technology at the same university. Njood employed as a collaborator at Imam Mohammad Ibn Saud Islamic University and has an interest in the academic and scientific community, particularly in the domains of AI, machine learning, and deep learning. She can be contacted at email: 441212541@qu.edu.sa.



**Amal A. Al-Shargabi**    is a associate professor at the College of Computer, Qassim University. She received her Ph.D. degree in Information Technology and Quantitative Sciences from Universiti Teknologi MARA (UiTM), Malaysia. She earned her Master's degree, with an award of excellence, in Computer Science from the same university. Her research interests include empirical program comprehension, empirical software engineering, and machine learning. She has attended and participated in various international conferences such as the IEEE//ACM International Conference on Program Comprehension, the International Conference on Soft Computing in Data Science, and the IEEE Conference on e-Learning, e-Management, and e-Services. She has received several awards, including the Invention, Innovation and Design Exhibition (IIDX'16) Award, Malaysia, and the Three Minute Thesis (3MT'16) Award, Malaysia. She can be contacted at email: amal.al-shargabi@dmu.ac.uk and a.alshargabi@qu.edu.sa.